INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Quasi-conjugate Bayes estimates for GPD parameters and application to heavy tails modelling

Jean Diebolt — Mhamed-Ali El-Aroui — Myriam Garrido — Stéphane Girard

N° 4803

Avril 2003

THÈME 4

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4803--FR+ENG

Quasi-conjugate Bayes estimates for GPD parameters and application to heavy tails modelling

Jean Diebolt , Mhamed-Ali El-Aroui , Myriam Garrido , Stéphane Girard

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet IS2

Rapport de recherche n° 4803 — Avril 2003 — 29 pages

Abstract: We present a quasi-conjugate Bayes approach for estimating Generalized Pareto Distribution (GPD) parameters, distribution tails and extreme quantiles within the Peaks-Over-Threshold framework. Damsleth conjugate Bayes structure on Gamma distributions is transfered to GPD. Bayes credibility intervals are defined, they provide assessment of the quality of the extreme events estimates. Posterior estimates are computed by Gibbs samplers with Hastings-Metropolis steps. Even if non-informative priors are used in this work, the suggested approach could incorporate informative priors, it brings solutions to the problem of estimating extreme events when data are scarce but expert opinion is available. It is shown that the obtained quasi-conjugate Bayes estimators compare well with the GPD standard estimators on simulated and real data sets.

Key-words: Extreme quantiles, Gamcon II distribution, Generalized Pareto Distribution, Gibbs Sampler, Peaks over thresholds (POT).

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 Montbonnot-St-Martin (France)
Téléphone : +33 4 76 61 52 00 —Télécopie +33 4 76 61 52 52

Estimateur bayésien quasi conjugué pour les paramètres de la loi GPD et application à la modélisation des queues de distribution lourdes

Résumé : Nous présentons une approche bayésienne quasi conjuguée pour l'estimation des paramètres de la loi de Pareto généralisée (GPD, *Generalized Pareto Distribution*), des queues de distribution et de quantiles extrêmes à l'aide de la méthode des excès. La structure bayésienne conjuguée proposée par Damsleth est tranférée aux lois GPD. Nous définissons des intervalles de crédibilité bayésiens produisant une information sur la qualité de nos estimations. Celles-ci sont calculées à partir de la loi a posteriori obtenue à l'aide d'un algorithme de Gibbs avec une étape de Hastings-Metropolis. Même si l'on utilise dans ce travail uniquement des lois a priori non informatives, l'approche suggérée peut incorporer des lois a priori informatives, ce qui apporte des solutions au problème de la modélisation d'événements rares lorsque l'on dispose de peu de données mais qu'un avis d'expert est disponible. Nous comparons les estimateurs bayésiens quasi conjugués obtenus à des estimateurs standards pour des données simulées et réelles.

Mots-clés : quantile extrême, loi gamcon II, loi de Pareto généralisée, algorithme de Gibbs, méthode des excès.

1 Introduction

Motivated by univariate tail and extreme quantile estimation, our goal is to develop new Bayesian procedures for making statistical inference on the shape and scale parameters of Generalized Pareto Distributions when used to model heavy tails and estimate extreme quantiles.

Most of the present methods of univariate tail estimation and statistical inference for extreme quantiles rely on a basic result: see Balkema and de Haan (1974) and Pickands (1975). Loosely speaking, this result states that if X is a random variable with distribution function (d.f.) F, then the conditional distribution of the excess $Y = X - u$ above the threshold u given that $X > u$ can in general be approximated by a properly scaled Generalized Pareto Distribution (GPD) as u tends to the endpoint of F. For u large enough, this result means that it is possible to approximate the upper tail $1 - F(u + y)$, $y \geq 0$, by $(1 - F(u))$ times a rescaled GPD. Given an n–sample and a sufficiently large threshold $u = u_n$, the shape and scale parameters of the approximating GPD are estimated on the basis of the excesses above u_n. The estimates are then usually plugged into the GPD d.f. and extreme quantile estimates are deduced. In this perspective, good estimation procedures for the shape and scale parameters of a GPD on the basis of approximately independent and identically distributed (i.i.d.) observations are necessary for accurate tail estimation. This is the basis of the Peaks Over Threshold (POT) method, described for example in Davison and Smith (1990) or in the monographs Embrechts, Klüppelberg and Mikosh (1997) and Reiss and Thomas (2001).

More precisely, let us assume that observations of a studied phenomenon $x_1, x_2, \ldots, x_n$ are issued from i.i.d. random variables $X_1, X_2, \ldots, X_n$ with unknown common d.f. F. Suppose that one needs to estimate either extreme quantiles q_{1-p} of F (i.e., $1 - F(q_{1-p}) = p$ with $p \in (0,\, 1/n]$ typically), or the extreme tail of F (i.e., $1 - F(x)$ for $x \geq x_{n,\,n}$, where $x_{1,\,n} \leq \ldots \leq x_{n,\,n}$ denote the ordered observations). It is usually recommended to use the POT method, where only observations x_i exceeding a sufficiently high threshold u_n are considered. In view of the theorem of Balkema and de Haan (1974) and Pickands (1975) the probability distribution of the $k = k_n$ positive excesses $y_j = x_{n-j+1,\,n} - u_n$ for $j = 1, \ldots, k$, where $x_{n-k,\,n} < u_n \leq x_{n-k+1,\,n}$, can be approximated for large u_n by a GPD$(\gamma,\, \sigma)$ distribution with scale parameter $\sigma > 0$ and shape parameter γ. The d.f. of GPD$(\gamma,\, \sigma)$ is

$$F_{\gamma,\,\sigma}(y) = \begin{cases} 1 - \left(1 + \dfrac{\gamma y}{\sigma}\right)_+^{-1/\gamma} & \text{if } \gamma \neq 0 \\[2ex] 1 - \exp\left(-\dfrac{\gamma y}{\sigma}\right) & \text{if } \gamma = 0\,, \end{cases} \tag{1}$$

with $y_+ = \max(y,\, 0)$, where $y \in \mathbb{R}^+$ when $\gamma \geq 0$, and $y \in [0,\, -\sigma/\gamma]$ when $\gamma < 0$.

Estimating the shape and scale parameters, γ and σ, is not easy. The maximum likelihood (ML) estimators (Smith, 1987) may be numerically hardly tractable, see Davison and Smith (1990) and Grimshaw (1993). Smith (1987) has shown that estimating GPD parameters with MLE's is a non regular problem for $\gamma < -1/2$. Moreover, the properties of MLE's meet their asymptotic theory only when the sample size (in our case the number k of exceedances) is larger than about 500. Alternative estimators have been proposed by Hosking and Wallis (1987): their linear estimators based on probability weighted moments (PWM) are easily computed and generally reasonably efficient when $-0.4 < \gamma < 0.4$ approximately (see, e.g., the comparative studies of Hosking and Wallis, 1987, and Singh and Guo, 1997). Castillo and Hadi (1997) have proposed other estimators based on the elemental percentile method (EPM), involving intensive computations. Their numerical simulations show that EPM estimators are more efficient than PWM ones only when $\gamma < 0$. The principle of maximum entropy has also been used (Singh and Guo, 1997): the authors conclude that their estimators are comparable in terms of bias and relative mean squared error to PWM. Finally, see Dupuis (1998) for a robust estimation procedure based on optimal bias-robust estimates.

Semiparametric estimators of γ along with related estimators of extreme quantiles have been intensively studied. For example, the Hill estimator presented by Hill (1975) and studied in Haeusler and Teugels (1985) and Beirlant and Teugels (1989), among many others. Two classic extensions of the Hill estimators are:

- The moment tail index estimator (denoted hereafter MTI(DEdH)) of Dekkers, Einmahl and de Haan (1989);
- The Zipf estimator, see Schultze and Steinebach (1996), and its generalization by Beirlant, Dierckx and Guillou (2001), denoted hereafter ZipfG.

Most of these semiparametric estimators do not perform much better than parametric ones when applied to sets of excesses. Only the ZipfG estimator seems to outperform the other ones.

Actually, bias-corrected semiparametric estimators of γ and extreme quantiles have recently been introduced and studied. See, e.g., Beirlant, Vynckier and Teugels (1996), Beirlant, Dierckx, Goegebeur and Matthys (2000), Beirlant, Dierckx, Guillou and Stărică (2002). These estimators, based on the second-order theory of regularly varying functions, are not studied in this paper.

The present paper presents a new Bayesian inference approach for GPD's with $\gamma > 0$. In a number of application areas such as structural reliability (see for example Grimshaw, 1993) and excess-of-loss reinsurance (see Reiss and Thomas, 2001), tail estimation based on small or moderate data sets is needed. In such situations Bayes procedures can be used to capture and take into account all available information including expert information even when it is loose. Moreover, in the realm of tail inference, evaluating the imprecision of estimates is of vital importance. Bootstrap methods have been suggested to assess this imprecision. But standard boostrap based on larger values of ordered samples is known to be inconsistent, whereas standard bootstrap based on excess samples has not second-order coverage accuracy and is imprecise when sample sizes are not extremely large (e.g., Bacro and Brito (1993) Caers, Beirlant and Vynckier, 1998). On the contrary, in the Bayesian context, credibility regions and marginal credibility intervals for GPD parameters and related high quantiles provide a non-asymptotic geometry of uncertainty directly based on outputs of the procedure, thus shortcutting bootstrap. Actually, imprecision measures making use of posterior distributions can loosely be seen as analogues of GPD-based parametric bootstrap. For all these reasons (expert information, credibility intervals) easily implementable Bayesian inference procedures for GPD's are highly desirable to study excess samples in the scope of POT methodology.

The restriction $\gamma > 0$ is not too damaging, since several major application areas are connected to heavy-tailed distributions. Our approach can also be tried for data issued from distributions suspected to lie in Gumbel's maximum domain of attraction where $\gamma = 0$. In the latter case, direct Bayesian analysis of the exponential distribution can be made in parallel (see Appendix A).

For other papers on Bayesian approaches to high quantile estimation, see, e.g., Coles and Tawn (1996), Coles and Powell (1996), Reiss and Thomas (1999), and the monographs Reiss and Thomas (2001) and Coles (2001) along with references therein.

Our starting point is a representation of heavy-tailed GPD's as mixtures of exponential distributions with a gamma mixing distribution. Since the Bayesian conjugate class for gamma distributions is documented (Damsleth, 1975) we only have to transfer it to GPD's. As described in Section 2, this provides a natural Bayesian quasi-conjugate class for heavy-tailed GPD's, leading to comparatively simple calculations and highly efficient computations through Gibbs sampling. Hierarchical structures can be built over this quasi-conjugate class: this will be discussed in a forthcoming work.

The computation of the posterior distributions is numerically implemented through an efficient Markov Chain Monte Carlo algorithm in Section 2. Section 3 compares our Bayes estimates to ML, PWM, moment tail index MTI(DEdH) and generalized Zipf (ZipfG) estimates on excess samples through Monte Carlo simulations. Section 4 is devoted to benchmark real data sets. Finally, Section 5 lists some conclusions and presents forthcoming research projects.

2 Bayesian inference for GPD parameters

The standard parameterization of heavy-tailed GPD distributions described by (1) when $\gamma > 0$ is now replaced by a more convenient one depending on the two positive parameters $\alpha = 1/\gamma$ and $\beta = \sigma/\gamma$. The re-parameterized version GPD(α, β) has d.f.

$$F_{\alpha,\beta}(y) = F(y \mid \alpha, \beta) = 1 - \left(1 + \frac{y}{\beta}\right)^{-\alpha}, \quad y \geq 0, \tag{2}$$

and probability density function (p.d.f.)

$$f(y \mid \alpha, \beta) = \frac{\alpha}{\beta}\left(1 + \frac{y}{\beta}\right)^{-\alpha-1}, \quad y \geq 0. \tag{3}$$

We assume that we have observations $\mathbf{y} = (y_1, \ldots, y_k)$ which are realizations of i.i.d. random variables $Y_1, \ldots, Y_k$ approximately issued from (2)–(3). Typically, they represent excesses above some sufficiently high threshold u. The latter means that for each $j \leq k$, there exists an integer $i \leq n$ such that $Y_j = X_i - u$, $X_i > u$, where the $X_i's$ are assumed i.i.d. and issued from a distribution in Fréchet's maximum domain of attraction: DA(Fréchet). Remark that the case where the common distribution of the $X_i's$ is in Gumbel's maximum domain of attraction (see Appendix A) can also be covered by considering the limiting situation $\alpha \to +\infty$ and $\beta \to +\infty$ with $\beta/\alpha \to \sigma > 0$.

Our starting point is the following mixture representation for (3), see Reiss and Thomas (2001) page 157:

$$f(y \,|\, \alpha, \beta) = \int_0^\infty z\, e^{-yz}\, g(z \,|\, \alpha, \beta)\, dz\,, \tag{4}$$

where for $z \geq 0$, $g(z|\,\alpha, \beta) = [\beta^\alpha / \Gamma(\alpha)]\, z^{\alpha-1} e^{-\beta z}$ is the density of the Gamma(α, β) distribution with shape and scale parameters α and β. The previous representation stands only in DA(Fréchet) as α and β have to be non-negative (as parameters of a gamma distribution) which implies $\gamma > 0$.

There is no Bayesian conjugate class for GPD's. Nevertheless, as shown below, the mixture form (4) allows us to make use of the conjugate class for gamma distributions to construct a suitable quasi-conjugate class for GPD's.

2.1 Conjugate distributions for Gamma(α, β)

According to Damsleth (1975), the description of the conjugate class for gamma distributions relies on the so-called type II Gamcon distributions: for $x > 0$, the density of the Gamcon II(c, d) distribution with parameters $c > 1$ and $d > 0$ is

$$\xi_{c,\,d}(x) = I_{c,\,d}^{-1}\, \Gamma(dx + 1)\, (\Gamma(x))^{-d}\, (c\, d)^{-dx}\,, \tag{5}$$

where $I_{c,\,d} = \int_0^\infty \Gamma(dx + 1)(\Gamma(x))^{-d}(c\,d)^{-dx}\, dx$. The density $\xi_{c,\,d}$ has roughly the shape of a gamma density, $\lim_{x \to 0+} \xi_{c,\,d}(x) = 0$ and $\xi_{c,\,d}$ is upper tail equivalent to a gamma density up to some multiplicative constant factor. Let $\mathbf{z} = (z_1, \ldots, z_k)$ denote a k-sample of realizations of i.i.d. random variables $Z_1, \ldots, Z_k$ issued from Gamma(α, β). According to Damsleth (1975), Theorem 2, the conjugate prior density on (α, β) with hyperparameters $\delta > 0$ and $\eta > \mu > 0$ is given by $\pi_{\delta,\,\eta,\mu}(\alpha, \beta) = \pi(\alpha)\, \pi(\beta \,|\, \alpha)$ where

- $\pi(\alpha)$ is the density of Gamcon II($c = \eta/\mu$, $d = \delta$) ;
- $\pi(\beta|\,\alpha)$ is the density of Gamma($\delta\alpha + 1$, $\delta\eta$).

The corresponding posterior densities are given by:

- The conditional density of α given $\mathbf{z}$, denoted $\pi(\alpha|\,\mathbf{z})$, is Gamcon II(η'/μ', δ') with

$$\delta' = \delta + k\,, \quad \eta' = \frac{\delta\eta + \sum_{i=1}^k z_i}{\delta + k} \quad \text{and} \quad \mu' = \mu^{\delta/(\delta+k)} \left(\prod_{i=1}^k z_i \right)^{1/(\delta+k)} \quad ; \tag{6}$$

- The conditional density of β given α and $\mathbf{z}$, denoted $\pi(\beta|\,\alpha,\,\mathbf{z})$, is Gamma$(\delta'\alpha + 1,\, \delta'\eta')$.

The hyperparameters η and μ act on the sufficient statistics $\sum_{i=1}^k z_i$ and $\sum_{i=1}^k \ln z_i$, whereas δ tunes the importance of these modifications. For $\delta = 1$, the introduction of these prior distributions can loosely be interpreted as artificially adding one observation $z_0 = \eta$ in the arithmetic mean, and another observation $z_0' = \mu$ in the geometric mean, with $z_0 / z_0' = \eta/\mu = c > 1$.

2.2 Transfer paradigm

Now the question is: How can we deduce the posterior density of the GPD parameters $(\alpha,\,\beta)$ from the gamma conjugate structure, starting with the prior density $\pi_{\delta,\,\eta,\mu}(\alpha,\,\beta)$?

At this point, we face the following general question. Let $\mathbf{y}$ and $\mathbf{z}$ be realizations of two random vectors $\mathbf{Y}$ and $\mathbf{Z}$ defined on the probability space $(I\!\!R^k,\, \mathcal{B}(I\!\!R^k))$. Consider a parametric family of densities for $\mathbf{Z}$, $\{g_\theta\}_{\theta\in\Theta}$ (also denoted $g(\bullet|\,\theta)$) where $\Theta \subset I\!\!R^p$. Assume that a natural parametric family of densities for $\mathbf{Y}$, $\{f_\theta\}_{\theta\in\Theta}$, can be deduced from $\{g_\theta\}_{\theta\in\Theta}$ as follows:

$$f_\theta(\mathbf{y}) = f\,(\mathbf{y}\,|\,\theta\,) \;=\; \int_{I\!\!R^k} p\,(\mathbf{y}\,|\,\mathbf{z}\,)\; g\,(\mathbf{z}\,|\,\theta\,)\; d\mathbf{z}\,, \tag{7}$$

where $\mathbf{y} \mapsto p(\mathbf{y}|\,\mathbf{z})$ is a probability density function for all $\mathbf{z} \in I\!\!R^k$ (i.e., $p(\mathbf{y}|\,\mathbf{z})$ is a transition density). Suppose that Bayesian inference is already documented for the family $\{g_\theta\}_{\theta\in\Theta}$. How can we transfer it to the family $\{f_\theta\}_{\theta\in\Theta}$? In our setting $\theta = (\alpha,\,\beta)$, f_θ is given by (3) and g_θ is the density of Gamma$(\alpha,\,\beta)$, as indicated above.

Notations and Assumptions –

1. The likelihood function of the observations $\mathbf{y}$ for f_θ writes $f(\mathbf{y}|\,\theta) = \prod_{i=1}^k f(y_i|\,\theta)$.
2. Similarly, the likelihood function of $\mathbf{z}$ for g_θ writes $g(\mathbf{z}|\,\theta) = \prod_{i=1}^k g(z_i|\,\theta)$.
3. We set $p(\mathbf{y}|\,\mathbf{z}) = \prod_{i=1}^k p(y_i|\,z_i)$.
4. We let $\pi(\theta|\,\mathbf{y})$ denote the posterior density of θ given the observations $\mathbf{y} = (y_1, \ldots, y_k)$ corresponding to f_θ and the prior density $\pi(\theta)$:
$$\pi\,(\theta\,|\,\mathbf{y}\,) \;=\; \frac{f\,(\mathbf{y}\,|\,\theta\,)\;\pi\,(\theta)}{f_\pi\,(\mathbf{y})}, \quad \text{where} \quad f_\pi(\mathbf{y}) \;=\; \int_\Theta f\,(\mathbf{y}|\,\theta')\,\pi(\theta')\,d\theta'\,. \tag{8}$$
5. Similarly, $\pi(\theta|\,\mathbf{z})$ denotes the posterior density of θ given $\mathbf{z} = (z_1, \ldots, z_k)$ corresponding to g_θ and the prior density π:
$$\pi\,(\theta\,|\,\mathbf{z}\,) \;=\; \frac{g\,(\mathbf{z}\,|\,\theta\,)\;\pi\,(\theta)}{g_\pi\,(\mathbf{z})}, \quad \text{where} \quad g_\pi(\mathbf{z}) \;=\; \int_\Theta g\,(\mathbf{z}|\,\theta')\,\pi\,(\theta')\;d\theta'\,. \tag{9}$$

6. We further denote

$$q_\pi(\mathbf{z}\,|\,\mathbf{y}) \;=\; \frac{p(\mathbf{y}\,|\,\mathbf{z})\,g_\pi(\mathbf{z})}{\int p(\mathbf{y}\,|\,\mathbf{z}')\,g_\pi(\mathbf{z}')\;d\mathbf{z}'}\,. \tag{10}$$

   For each $\mathbf{y}$, the function $\mathbf{z} \mapsto q_\pi(\mathbf{z}|\,\mathbf{y})$ is a p.d.f. with respect to Lebesgue measure: $q_\pi(\mathbf{z}|\,\mathbf{y})$ is a transition density.

The following straightforward result shows that the posterior distribution of θ given $\mathbf{y}$ is a mixture of the posterior distributions of θ given $\mathbf{z}$ with mixing density $q_\pi(\mathbf{z}|\,\mathbf{y})$. See Garrido (2002) for more details.

Proposition 1 *Under the previous assumptions we have*

$$\pi(\theta\,|\,\mathbf{y}) \;=\; \int q_\pi(\mathbf{z}\,|\,\mathbf{y})\;\pi(\theta\,|\,\mathbf{z})\;d\mathbf{z}\,, \tag{11}$$

where q_π is defined by (10).

REMARK 1 . – It follows from Proposition 1 that each posterior moment given $\mathbf{y}$ is the integral of the corresponding posterior moment given $\mathbf{z}$ with respect to $q_\pi(\mathbf{z}|\,\mathbf{y})$.

REMARK 2 . – In the GPD setting, we have

$$g(\mathbf{z}\,|\,\theta) = \frac{\beta^{k\alpha}}{(\Gamma(\alpha))^k}\left(\prod_{i=1}^{k} z_i\right)^{\alpha-1} e^{-\beta\sum_{i=1}^{k} z_i} \quad \text{and} \quad p(\mathbf{y}\,|\,\mathbf{z}) = \left(\prod_{i=1}^{k} z_i\right)\exp\left(-\sum_{i=1}^{k} y_i z_i\right).$$

REMARK 3 . – Let $f_{\text{pred}}(y_0\,|\,\mathbf{y}) = \int f(y_0\,|\,\theta)\;\pi(\theta\,|\,\mathbf{y})\;d\theta$ be the posterior predictive density based on $\mathbf{y}$, and $g_{\text{pred}}(z_0\,|\,\mathbf{z}) = \int g(z_0\,|\,\theta)\;\pi(\theta\,|\,\mathbf{z})\;d\theta$ be the posterior predictive density based on $\mathbf{z}$. These predictive densities are linked by

$$f_{\text{pred}}(y_0\,|\,\mathbf{y}) \;=\; \int q_\pi(\mathbf{z}\,|\,\mathbf{y})\left(\int p(y_0\,|\,z_0)\,g_{\text{pred}}(z_0\,|\,\mathbf{z})\;dz_0\right)d\mathbf{z}\,. \tag{12}$$

Unfortunately, the functions g_π (see (9)–(10)) and $\mathbf{z} \mapsto q_\pi(\mathbf{z}|\,\mathbf{y})$ (see (10)–(11)) are not expressible in analytical close form. Therefore, a numerical algorithm is needed. The mixture representation (11) allows us to design a simple and efficient Gibbs sampler. It is presented in the next Subsection.

2.3 Gibbs sampling

We consider again the GPD setting where $\mathbf{y}$ denotes a k-sample from GPD(α, β). A Gibbs sampler is used to get approximate simulations from the posterior density of θ given $\mathbf{y}$,

$\pi(\theta|\mathbf{y})$. Damsleth's priors are used for $\theta = (\alpha, \beta)$. The proposed sampler generates a Markov chain whose equilibrium density (denoted $\pi(\mathbf{z}, \theta|\mathbf{y})$) is the joint density of $(\mathbf{Z}, \theta)$ conditionally on $\mathbf{Y} = \mathbf{y}$ where $\mathbf{Y}$ and $\mathbf{Z}$ denote respectively random vectors of k independent Gamma(α, β) and GPD(α, β) random variables. To implement the Gibbs sampler we first note that within the general setting of subsection 2.2, the conditional density of θ given $(\mathbf{y}, \mathbf{z})$ is independent of $\mathbf{y}$: $\pi(\theta|\mathbf{y}, \mathbf{z}) = \pi(\theta|\mathbf{z})$, and the conditional density of $\mathbf{Z}$ given $(\mathbf{y}, \theta)$ is

$$f(\mathbf{z} \mid \mathbf{y}, \theta) = \frac{p(\mathbf{y} \mid \mathbf{z})\, g(\mathbf{z} \mid \theta)}{f(\mathbf{y} \mid \theta)}. \tag{13}$$

In our GPD setting,

$$f(\mathbf{z} \mid \mathbf{y}, \theta) = \prod_{i=1}^{k} f(z_i \mid y_i, \theta) \propto \prod_{i=1}^{k} z_i^{\alpha}\, e^{-(\beta + y_i) z_i}\, \mathbb{1}_{z_i > 0}. \tag{14}$$

It follows that for $i \leq k$, conditionally on θ and $Y_i = y_i$, $Z_i \sim \text{Gamma}(\alpha + 1, \beta + y_i)$ independently.

This yields the following intertwining sampler, where simulation from the posterior density of θ given the observations $\mathbf{y}$ is based on successive simulations alternatively from the conditional density of $\mathbf{Z}$ given $(\mathbf{y}, \theta)$ and the posterior density of θ given $\mathbf{z}$. Let $\theta^{(m)}$ denote the current parameter value at iteration m. The next iteration:

1. `independently simulates each` $z_i^{(m+1)}$ `from Gamma`$(\alpha^{(m)} + 1, \beta^{(m)} + y_i)$ `;`
2. `simulates` $\theta^{(m+1)}$ `from` $\pi(\theta|\,\mathbf{z}^{(m+1)})$.

In such a setting, both $(\mathbf{z}^{(m)})_{m \geq 0}$ and $(\theta^{(m)})_{m \geq 0}$ are Markov chains. The simulation step of $\theta^{(m+1)}$ is split into the marginal simulation of $\alpha^{(m+1)}$ and the conditional simulation of $\beta^{(m+1)}$ given $\alpha^{(m+1)}$. Finally, the iteration $m + 1$ of our Gibbs sampler:

1. `independently simulates each` $z_i^{(m+1)}$ `from Gamma`$(\alpha^{(m)} + 1, \beta^{(m)} + y_i)$ `;`
2. `simulates` $\alpha^{(m+1)}$ `from` $\pi(\alpha|\,\mathbf{z}^{(m+1)})$`, i.e. from Gamcon II`$(\eta'/\mu', \delta')$ `with` δ', η' `and` μ' `computed from` $\mathbf{z}^{(m+1)}$ `using equation (6) ;`
3. `simulates` $\beta^{(m+1)}$ `from Gamma`$(\delta'\alpha^{(m+1)} + 1, \delta'\eta')$.

When the equilibrium regime is nearly reached, simulated values of θ are approximatively issued from the posterior distribution of θ given $\mathbf{y}$.

Implementing the previous algorithm requires simulating Gamcon II distributions and choosing adequate values of the hyperparameters δ, η and μ of the priors $\pi(\alpha)$ and $\pi(\beta|\,\alpha)$.

2.4 Simulating Gamcon II distributions

Our sampling scheme involves simulations from Gamcon II$(c,\, d)$ distributions with $c = c' = \eta'/\mu'$, where η' and μ' are given by (6), and moderate to large values of $d = d' = \delta' = \delta + k$. Up to our knowledge, there is no standard algorithm for simulating such distributions. The simulation method that we present is based on a normal approximation to Gamcon II distributions using Laplace's method (this approximation is deeply studied in Garrido 2002).

Gamcon II$(c,\, d)$ can be approximated by a normal distribution having the same mode. It is proved in Garrido (2002) that this mode, $M_{c,\,d}$, is the unique root of the equation

$$\psi\,(d\,M_{c,\,d} + 1) - \psi\,(M_{c,\,d}) - \ln d - \ln c = 0\,, \tag{15}$$

where ψ denotes the digamma function: the derivative of the logarithm of $\Gamma(t)$,

$$\psi(t) = \frac{d}{dt} \ln \Gamma(t) = \frac{\Gamma'(t)}{\Gamma(t)}\,.$$

The standard deviation $S_{c,\,d}$ of the normal approximant distribution is computed through a Taylor expansion of the Gamcon II$(c,\, d)$ density in a neighborhood of its mode:

$$S_{c,\,d} = \frac{1}{\sqrt{d\psi'\,(M_{c,\,d}) - d^2\psi'\,(dM_{c,\,d} + 1)}}\,. \tag{16}$$

Garrido (2002) has established that $(1 - 1/d)/(\ln c + \ln d/2) \leq M_{c,\,d} \leq 2/\ln c$. In practice, $M_{c,\,d}$ is numerically approximated through the bisection method.

At each iteration, we simulate Gamcon II$(c',\, d')$ with the help of the independent Hastings-Metropolis algorithm, which requires a suitable proposal density. Actually, it is enough to make only one step of Hastings-Metropolis at each iteration of the Gibbs sampler: see Robert (1998). The proposal density must be as close as possible to the simulated density, Gamcon II$(c',\, d')$, and have heavier tails to ensure good mixing. Since Gamcon II densities have gamma-like tails (see the comments after (5) in Subsection 2.1 above), we cannot directly take the normal approximant density as a proposal. Rather, we have chosen a Cauchy proposal density as close as possible to the normal approximant density to Gamcon II$(c',\, d')$, i.e. with the same mode and modal value. Therefore at each iteration our Hastings-Metropolis step:

1. independently simulates a new Y from the Cauchy distribution with mode $M_{c',\,d'}$ and modal value $1/(S_{c',\,d'}\sqrt{2\pi})$;
2. computes the ratio

$$\rho = \min\left[1,\, \frac{f^\star_{\text{cauchy}}\left(\alpha^{(m)}\right)\, \xi_{c',\,d'}(Y)}{f^\star_{\text{cauchy}}(Y)\, \xi_{c',\,d'}\left(\alpha^{(m)}\right)}\right],$$

   where $f^\star_{\text{cauchy}}$ denotes the density of Y ;

3. `takes` $\alpha^{(m+1)} = Y$ `with probability` ρ `and` $\alpha^{(m+1)} = \alpha^{(m)}$ `otherwise.`

Since all the transition densities involved are positive, the resulting Markov chain $(\theta^{(m)})_{m\geq 0}$ is ergodic with unique invariant probability measure equal to the posterior distribution of θ given $\mathbf{y}$. Intensive numerical experiments reported in Garrido (2002) show that for $\delta > 0.5$, this Gibbs sampler with one Hastings-Metropolis step at each iteration converges quickly to its invariant distribution. Actually, it seems that discarding the first 500 iterations is sufficient, although refined MCMC control methods (for e.g. Chauveau and Diebolt, 1999) suggest a slightly more expensive strategy. For $\delta \leq 0.5$, we observed numerical instabilities.

Bayesian inference on the GPD parameters α, β is based on the outputs of this algorithm when it has approximately reached its stationary regime. We then record a sufficient number of realizations $\alpha^{(m)}$ and $\beta^{(m)}$.

2.5 Hyperparameters values

In this paper we suppose that no expert information is available for the choice of priors on the GPD parameters (α, β). Introduction of expert information will be discussed in a next paper. Here we take an empirical Bayes approach, where loose prior information is obtained from the available sample. The more natural approach to compute hyperparameter values of the priors $\pi(\alpha)$ and $\pi(\beta|\,\alpha)$ is to equate some of their location parameters (mean, median or mode) to frequentist estimates of α and β, denoted $\widehat{\alpha}$ and $\widehat{\beta}$ in this Subsection. Actually, we made use of the estimators $\widehat{\alpha}$ and $\widehat{\beta}$ induced by the Hill procedure, since they are easily computed and always provide positive values of $\widehat{\alpha}$ and $\widehat{\beta}$, i.e. of $\widehat{\gamma}$ as required in our procedure.

We took $\delta = 1$ both for convenience (for $\delta = 1$, the prior $\pi(\alpha)$ reduces to a gamma distribution, see below) and because a small value of $\delta > 0$ indicates little confidence in prior information (see the comments after (5)): here, we adhere to a quasi non-informative approach. Recall that we observed numerical instabilities of our Gibbs sampler for $\delta < 0.5$.

Prior means were used to determine values of the hyperparameters η and μ: the mean of the prior distribution $\pi(\beta|\,\alpha)$ is $(\alpha+1)/\eta$. Taking this mean equal to $\widehat{\beta} = x_{n-k,\,n}$, the estimate induced by the Hill procedure, and replacing α by its Hill estimate

$$\widehat{\alpha} \;=\; k_n \left(\sum_{i=1}^{k_n} \ln x_{n-k_n+i,\,n} \;-\; \ln x_{n-k_n\,,\,n} \right)^{-1}$$

yields $\eta = (\widehat{\alpha}+1)/\widehat{\beta}$. When $\delta = 1$, the prior $\pi(\alpha)$ reduces to Gamma$(2, \ln(\eta/\mu))$. Its mean is $2/\ln(\eta/\mu)$. Solving the equation where this prior mean is set equal to $\widehat{\alpha}$ yields

$$\mu \;=\; \frac{\widehat{\alpha}\,+\,1}{\widehat{\beta}}\;\exp\left(-\,\frac{2}{\widehat{\alpha}}\right).$$

If prior modes are used instead of prior means, a similar approach leads to slightly different formulas for η and μ. Actually, with prior modes explicit expressions can be obtained for all $\delta > 0$. However, preliminary numerical experiments yielded better estimates with prior means.

2.6 Computation of Bayesian estimates

When the Gibbs sampler approximately reaches its stationary regime, K values denoted $(\alpha^{(m)}, \beta^{(m)})$, $m = 1, \dots, K$, are saved. This simulated sample is used to compute posterior means, medians or modes and estimate $(\alpha,\, \beta)$, leading to Bayesian estimates for $(\gamma,\, \sigma)$ and q_{1-p}. Concerning the estimation of an extreme quantile q_{1-p} ($\mathbf{y} = (y_1, \dots, y_k)$ is a sample of excesses over a threshold u), a sample of values $\widehat{q}_{1-p}^{(m)}$ is computed using POT from the simulated $(\alpha^{(m)}, \beta^{(m)})$'s:

$$\widehat{q}_{1-p}^{(m)} \;=\; u \;+\; \beta^{(m)} \left[\left(\frac{np}{k} \right)^{-1/\alpha^{(m)}} - 1 \right] . \tag{17}$$

Means, histograms and credibility intervals can then be computed and represented from (17). See sections 3 and 4 below.

REMARK 4 . – The posterior modes are more difficult to compute since one has first to construct smooth estimates of the joint posterior density of α and β. Numerical experiments reported in Garrido (2002) for both GPD generated data and excess samples led us to keep only posterior medians. □

Bayesian credibility intervals for α, β and q_{1-p} are obtained by sorting the corresponding simulated values obtained from the Gibbs sampler. Similarly, predictive quantile functions can be approximated through

$$\widehat{F^{-1}}_{\text{pred}}(y) \approx \frac{1}{K} \sum_{m=1}^{K} F^{-1}_{\alpha^{(m)},\, \beta^{(m)}}(y) \, . \tag{18}$$

The probability distribution of the observed sample $\mathbf{y}$ can be estimated either by GPD$(\widehat{\alpha},\, \widehat{\beta})$, or by the posterior predictive distribution. According to Reiss and Thomas (2001) page 157, a mixture of Pareto distributions with respect to a gamma mixing density is a log-Pareto distribution, which has a "super heavy" tail, i.e. the right tail of the survival function is of the order of $O((\ln x)^{-\nu})$ for some positive ν as $x \to \infty$. Although in our case posterior predictive distributions are not expressible in close form, their mixture form (see (11)–(12) and Remark 2) and the properties of Gamcon II distributions suggest that our predictive distributions have super heavy tails, too. This is comparable to the results of Appendix A, and leads us to discard estimation procedures involving posterior predictive distributions.

3 Comparative Monte Carlo simulations

Intensive Monte Carlo simulations were used to compare our Bayes quasi-conjugate estimators (denoted hereafter Bayes-QC) of γ and q_{1-p} with their counterparts when usual estimators of GPD parameters are used: Maximum Likelihood (ML), Moment Tail Index estimator (MTI(DEdH)), Generalized Zipf estimator (ZipfG) and Probability Weighted Moments (PWM).

3.1 The simulation design

Three probability distributions in DA(Fréchet) were considered in order to produce various excess samples:

- The Fréchet(1) distribution, for which $\gamma = 1$ and the second-order regular variation parameter (presented below) $\rho = -1$. The d.f. of Fréchet(β) for $\beta > 0$ is $F(x) = \exp\left(-x^{-1/\beta}\right)$, $x > 0$.
- The Burr$(1, 0.5, 2)$ distribution, for which $\gamma = 1$ and $\rho = -0.5$. The d.f. of Burr(β, τ, λ) for $\beta > 0$, $\tau > 0$, $\lambda > 0$ is $F(x) = 1 - [\beta/(\beta + x^\tau)]^\lambda$, $x > 0$.
- The Log-Gamma(2) distribution, for which $\gamma = 1$ and $\rho = 0$. A random variable Z is Log-Gamma(2) distributed when $\ln Z$ is Gamma(1, 2) distributed. The density of Log-Gamma(2) is $f(x) = x^{-2}(\ln x)^{-1}$, $x > 0$.

For each distribution, the second-order regular variation parameter ρ ($\rho \leq 0$) indicates the quality of approximation of F_u by a GPD(γ, $\sigma(u)$) for high values of u and suitable $\sigma(u)$'s. High values of $|\rho|$ indicate excellent fitting, whereas values of $|\rho|$ close to 0 indicate bad fitting. For the Fréchet(1) distribution, $\rho = -1$ reflects the fact that

$$1 - F(x) \sim 1 - \exp\left(-\frac{1}{x}\right) = \frac{1}{x}\left(1 - \frac{1}{2x}\,(1 + o(1))\right) \qquad \text{as } x \to +\infty\,.$$

For the Burr(1, 0.5, 2) distribution, $\rho = -0.5$ corresponds to the expansion

$$1 - F(x) \sim \frac{1}{\left(1 + x^{1/2}\right)^2} = \frac{1}{x}\left(1 - \frac{2}{\sqrt{x}}\,(1 + o(1))\right) \qquad \text{as } x \to +\infty\,.$$

Finally, for the Log-Gamma(2) distribution, $\rho = 0$ corresponds to

$$1 - F(x) \sim \frac{1}{x\,\ln x} \qquad \text{as } x \to +\infty\,.$$

For each one of these three probability distributions, 100 data sets of size $n = 500$ were independently generated. For each simulated data set and each value of $k = 5, 10, \ldots, 490, 495$, we performed 1 000 iterations of the Gibbs sampler and only the last 500 ones were kept.

3.2 First results

For each simulated data set and each value of $k = 5, 10, \ldots, 490, 495$ Bayes-QC estimates of γ and q_{1-p}, $p = 1/5\,000$, are computed as the medians of the resulting 500 $\widehat{\gamma}^{(m)}$'s and $\widehat{q}_{1-p}^{(m)}$'s given in the 500 last iterations of the Gibbs sampler. Figures 1–3 display the averages over the 100 data sets of these Bayesian estimates of γ and q_{1-p} as functions of k. The means and modes were also computed and gave quite similar results. They are not displayed here.

The ML, MTI(DEdH) and ZipfG estimates of γ and q_{1-p}, $p = 1/5\,000$, were also computed for each of those simulated data sets and the same values of k. Figures 1–3 display the averages over the 100 data sets of these estimates of γ and q_{1-p}.

The left panels of Figures 1–3 show that our estimates of γ based on posterior medians (continuous line curves) perform rather well compared to the other ones, and give estimates close to ML and ZipfG. The right panels of Figures 1–3 show that our estimates of q_{1-p} are very close to those obtained by ML. Both give better estimates than MTI(DEdH) but in general do not perform as well as ZipfG, although they seem to be more stable as functions of k. Again, credibility intervals for γ and q_{1-p} and potential improvements when prior information is available are strong arguments supporting the use of our Bayesian procedure.

3.3 Bayes credibility and Monte-Carlo confidence intervals

Monte Carlo simulations were also used to study whether Bayesian credibility intervals could be used as approximate frequentist confidence intervals for γ and q_{1-p}, in the present quasi non-informative approach. For each simulated data set, the last 500 iterations of our Gibbs sampler provide 500 estimates $\widehat{\gamma}^{(m)}$ and 500 estimates $\widehat{q}_{1-p}^{(m)}$, $m = 501, \ldots, 1\,000$. They can be sorted to provide approximate 90 % credibility intervals for γ and q_{1-p}. The precision of these credibility intervals was studied through Monte Carlo simulations: for each one of the three probability distributions considered and for each value of k, we counted the number of simulated data sets (out of 100) for which the true values of γ and q_{1-p} fell within the corresponding 90 % credibility intervals. Figure 4 exhibits the coverage rates for each simulated distribution and for $k = 5, 10, \ldots, 490, 495$. These credibility intervals are very accurate for the Fréchet distribution. For the Log-Gamma distribution, the credibility intervals have good coverage rates for q_{1-p} but not for γ.

This rather unexpected behavior when $\rho = 0$ can be explained in terms of penultimate approximation, see Diebolt, Guillou and Worms (2002): it can be proved that for $\rho = 0$, the distribution of excesses is better approximated by a GPD with scale parameter $\gamma + a_k(F)$, where $a_k(F)$ is some correction term, than by a GPD with scale parameter γ (see, e.g., the paper coauthored by Kaufmann, pages 183–190 in Reiss and Thomas (2001) along with references therein and Worms (2001)). This explains why in the case $\rho = 0$ the estimates of γ strongly deviate from the true value. Furthermore, Diebolt, Guillou and Worms (2002) have established for all sufficiently regular estimators $(\widehat{\gamma}, \widehat{\sigma})$ of the parameters (γ, σ) such as ML or PWM, that when $\rho = 0$ the estimated survival function $\bar{F}_{\widehat{\gamma},\widehat{\sigma}}$ is a bias-corrected

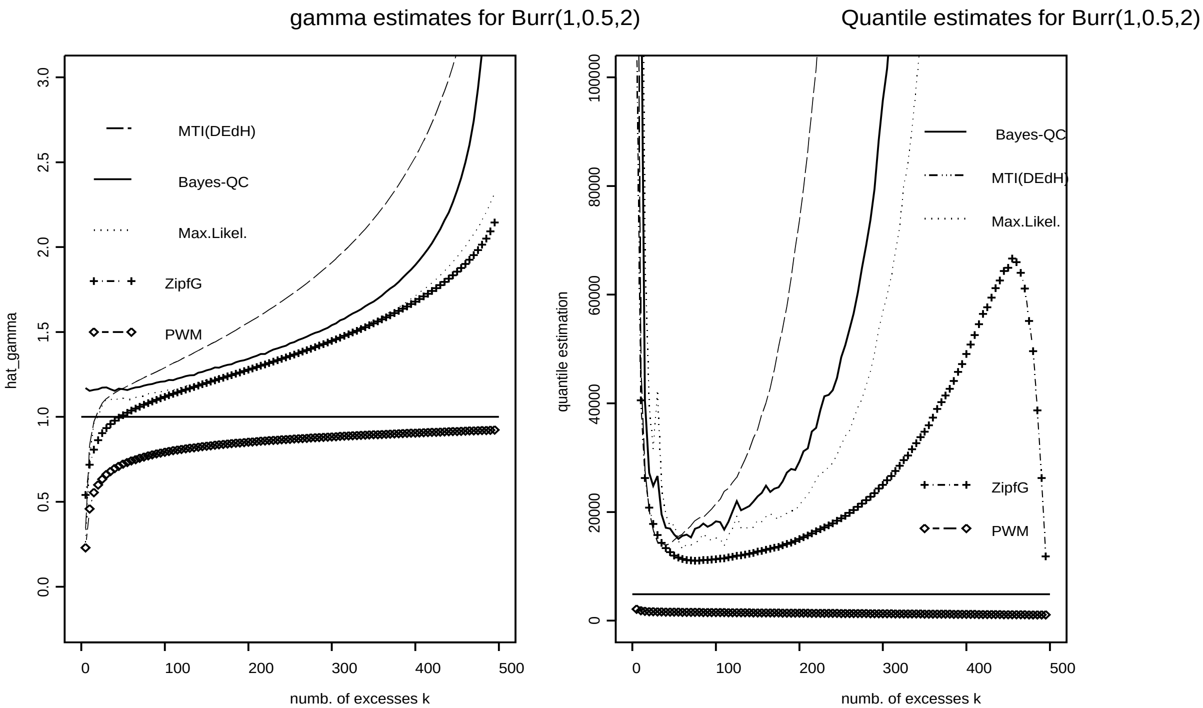

Figure 1: Means of $\hat{\gamma}$ and $\hat{q}$ on the 100 Monte Carlo replications from Burr(1, 0.5, 2) for different values of the number k of excesses.

estimate of $\bar{F}_{\gamma,\sigma}$. We think that this result partially explains the good and stable coverage rates observed for q_{1-p} for the Log-Gamma distribution.

Finally, Figure 1 suggests that the optimal value of k in the Burr(1, 0.5, 2) case is close to 90. For $k = 90$, the credibility intervals for both γ and q_{1-p} are still satisfactory.

These trials show that the credibility intervals computed through our procedure give very promising results.

For each one of the three simulated probability distributions and each value of $k \in \{5, 10, \ldots, 495\}$ the 100 simulated data sets give 100 estimates of (γ, q_{1-p}). The empirical 0.05 and 0.95 quantiles of the previous estimates give 90% Monte-Carlo Confidence intervals (MCCI) for (γ, q_{1-p}). The width of these MCCI's are used in Figure 5 to compare the precision of our Bayes Quasi-conjugate q_{1-p} estimator to ML, ZipfG and MTI(DEdH) estimators. For Burr data sets (left panel of Figure 5) ZipfG gives the most precise quantile estimators, it is followed by ML, Bayes-QC and MTI(DEdH). For Fréchet, Bayes-QC and ML have the

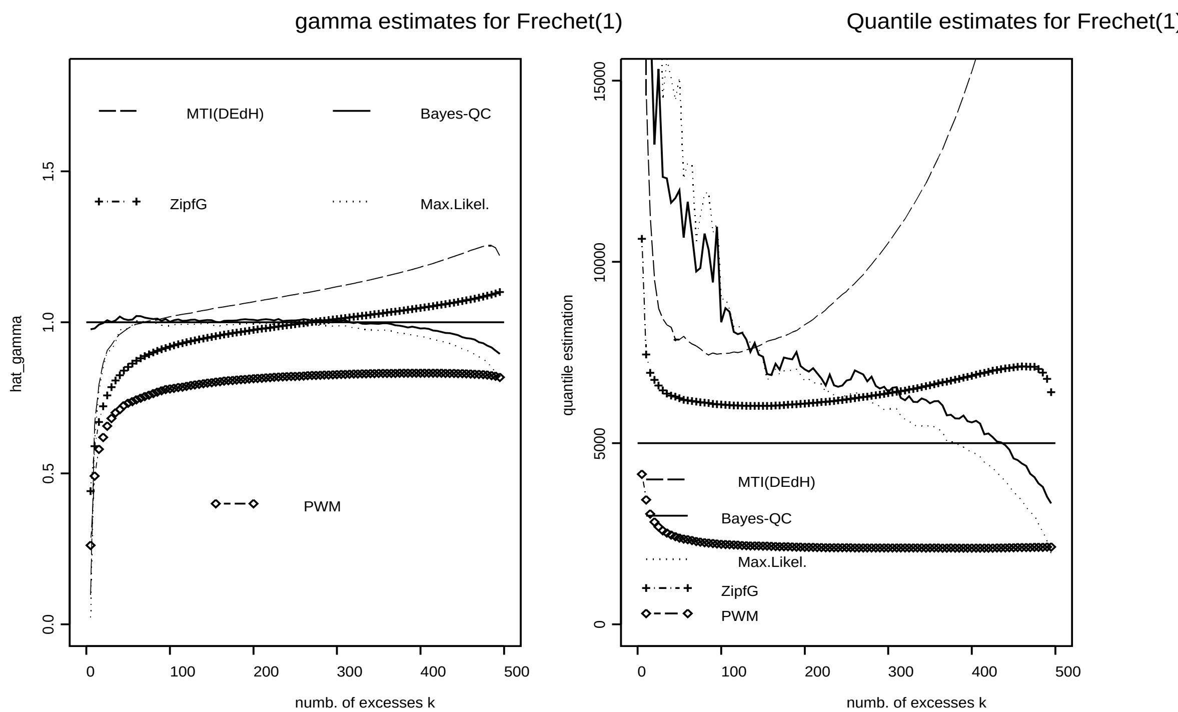

Figure 2: Means of $\hat{\gamma}$ and $\hat{q}$ on the 100 Monte Carlo replications from Fréchet(1) for different values of k.

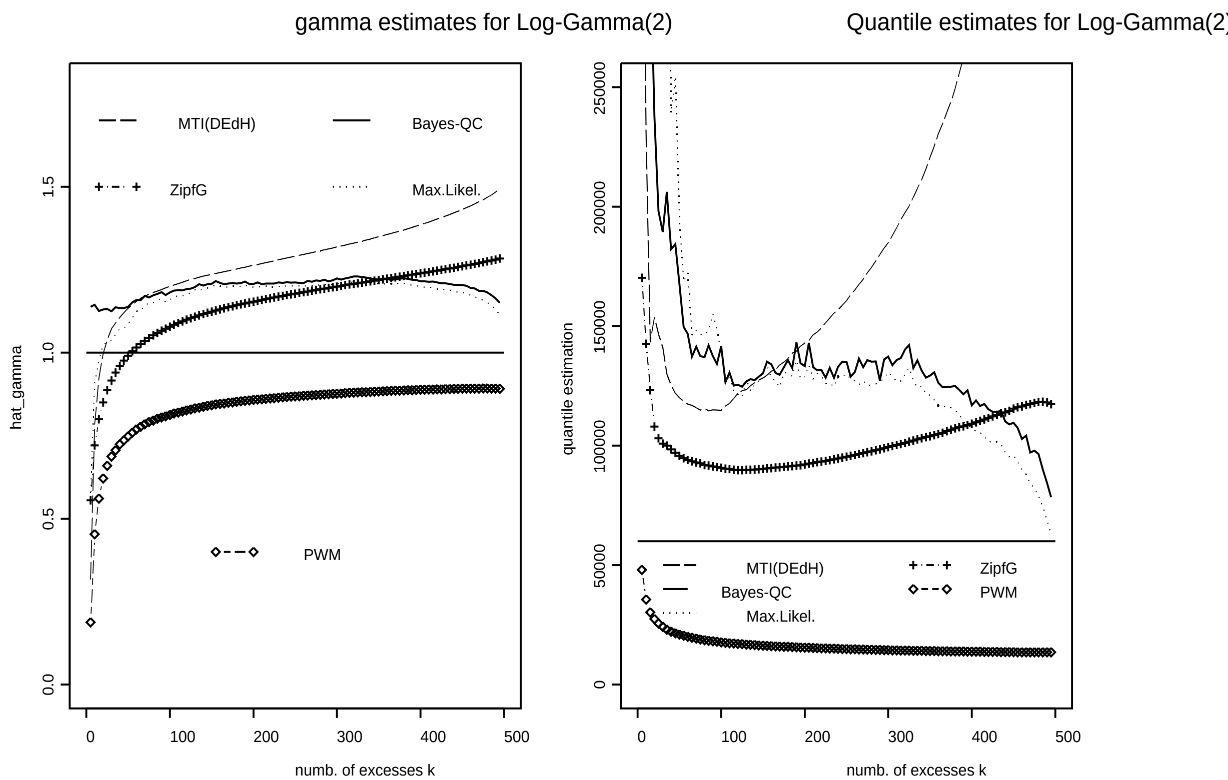

Figure 3: Means of $\hat{\gamma}$ and $\hat{q}$ on the 100 Monte Carlo replications from Log-Gamma(2) for different values of k.

best precisions. It is worth noting that Bayes-QC is used here with non-informative priors, its precision will increase when informative priors are used.

In Figure 6 the average Bayes credibility intervals (averaged over the 100 simulated data sets) are compared to the 90% Monte-Carlo Confidence intervals (MCCI) for our Bayes-QC estimator. For both Burr (left panel) and Fréchet (right panel) ditributions lower bounds of the average Bayes credibility intervals and MCCI are very close. Upper bounds of average Bayes credibility intervals are larger than those of MCCI. It is interesting to note that the width of the average Bayes credibility intervals are the narrowest for k where Bayes estimate of q_{1-p} is the closest to the true value q_{1-p}. This could be used to chose optimal values of the number of excesses k.

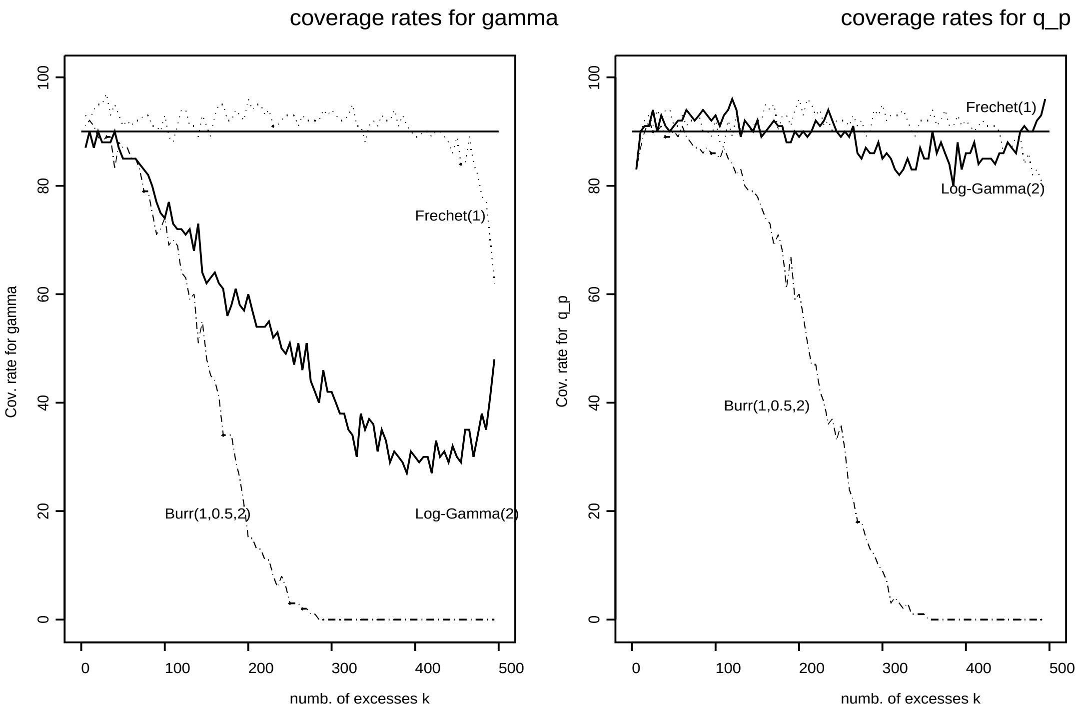

Figure 4: Coverage rates of our 90 % Bayesian credibility intervals for γ and q_{1-p} based on the 100 Monte Carlo replications for different values of k.

4 Application to real data sets

Here, advantages and good performance of our Bayesian estimators are illustrated through the analysis of extreme events described by two benchmark real data sets:

- Nidd river data. They are standard data in extreme value studies (used for example in Hosking and Wallis, 1987 and Davison and Smith, 1990). The raw data consist in 154 exceedances of the level 65 m^3s^{-1} by the river Nidd (Yorkshire, England) during the period 1934-1969 (35 years). The N-year return level is the water level which is exceeded on average once in N years. Hydrologists need to estimate extreme quantiles in order to predict return levels over long periods (250 years, i.e. $p = 9\,10^{-4}$, or even 500 years).

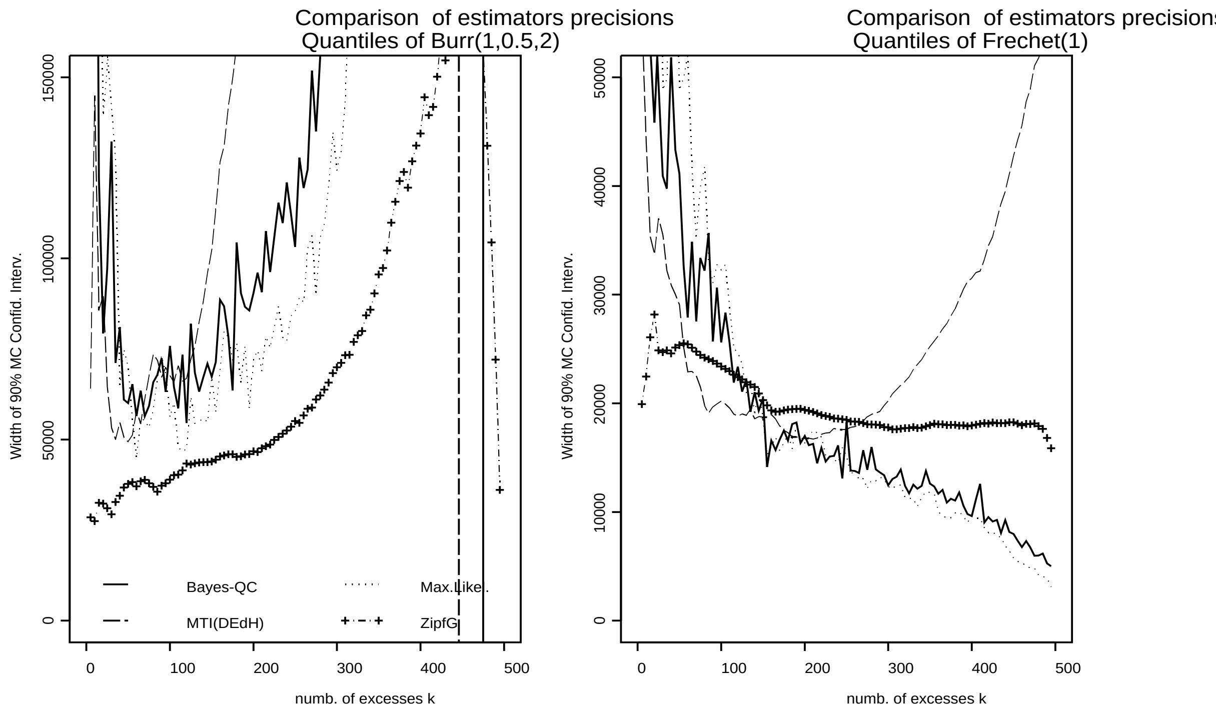

Figure 5: Width of 90% Monte-Carlo confidence intervals of $\hat{q}_{1-p}$ for different values of k.

– Fire reinsurance data. These data were first studied by Schnieper (1993) and Reiss and Thomas (1999). They represent insurance claims exceeding $u = 22$ millions of Norwegian Kröner from 1983 to 1992.

4.1 Nidd river data

Bayes quasi-conjugate estimates and related 90 % credibility intervals for γ and q_{1-p} are depicted in Figure 7 for several values of k. Compared to other estimators, our approach provides the most stable estimates as k varies. For 8 values of k Grimshaw's algorithm for computing ML estimates did not converge: see the broken ML curves in Figure 7. Histograms of γ's and q_{1-p}'s for $k = 82$ are displayed in Figure 8. Table 1 summarizes results of the estimation of the 50-year and 100-year return levels of the Nidd river when the threshold u is set equal to either 100 ($k = 39$) or 120 ($k = 24$).

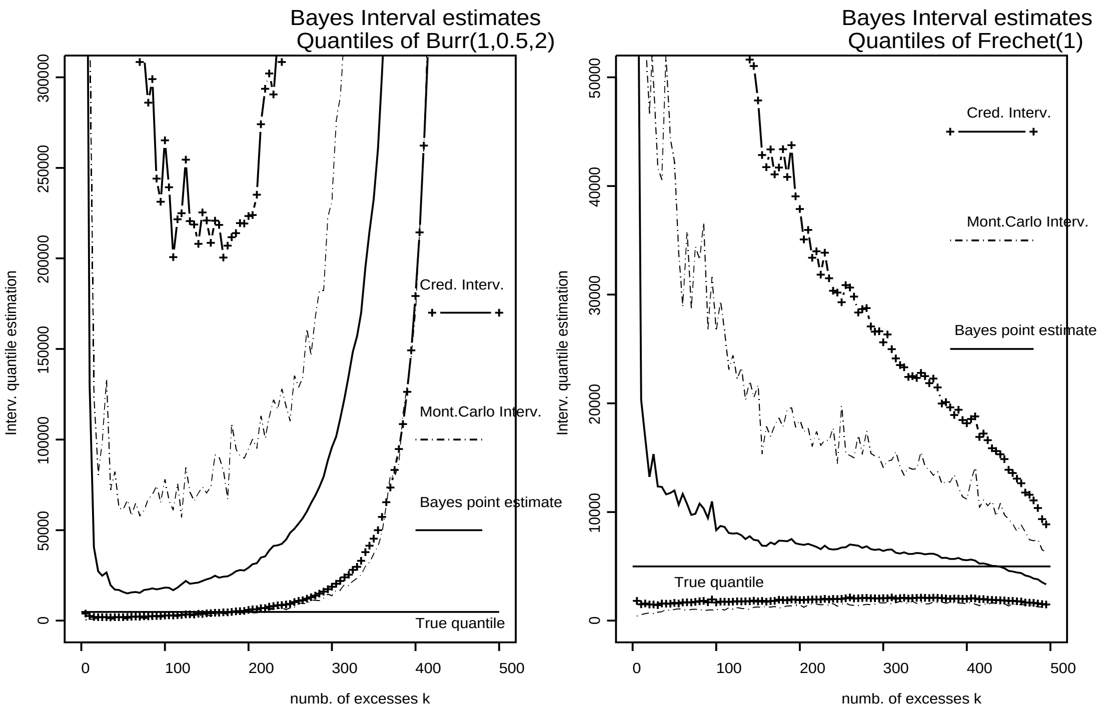

Figure 6: 90 % Bayes credibility and Monte-Carlo confidence intervals for q_{1-p} for different values of k.

REMARK 5 . – Recall that the N-year return level RL_N is the water level which is exceeded on average once in N years. Equations relating q_{1-p} to the return level RL_N follow from Davison and Smith (1990): $\widehat{RL}_N \simeq u - (\widehat{\sigma}/\widehat{\gamma})[1-(\widehat{\lambda}N)^{\widehat{\gamma}}]$, where it is assumed that the exceedance process is Poisson with annual rate λ. If we have observed k excesses above the threshold u during 35 years, then λ is estimated by $k/35$. It follows that $\widehat{RL}_N = \widehat{q}_{1-p}$ with $p \simeq k \,/\, \widehat{\lambda}Nn$. For the Nidd river data, this yields $p \simeq 35 \,/\, Nn$. Therefore, estimating the 100-year return level is equivalent to estimating q_{1-p} by $p = 35/(100 \times 154) \simeq 2.27\, 10^{-3}$. $\square$

As shown in Table 1, our credibility intervals with level 95 % (see the last column) compare well to the Bayesian confidence intervals obtained by Davison and Smith (1990), which are based on uniform priors for q_{1-p}, λ and γ (see Smith and Naylor, 1987 for more details). Actually, ours are slightly narrower.

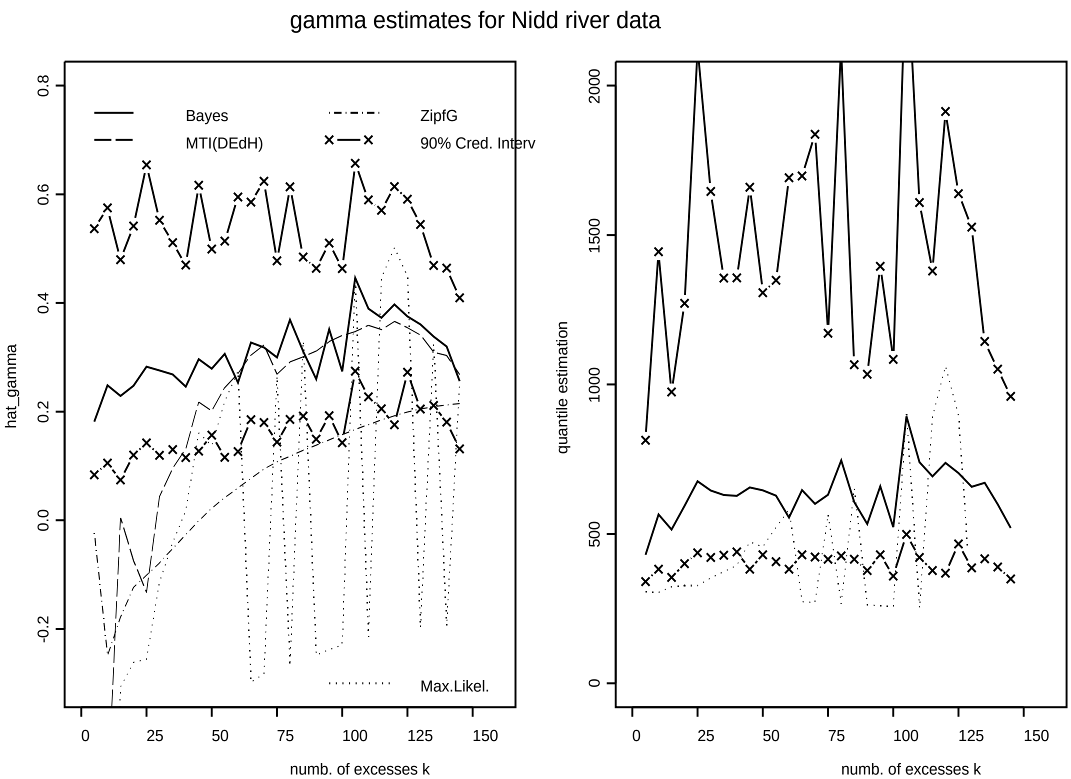

Figure 7: $\widehat{\gamma}$ and $\widehat{q}_{1-p}$ for the Nidd river data set with several values of k.

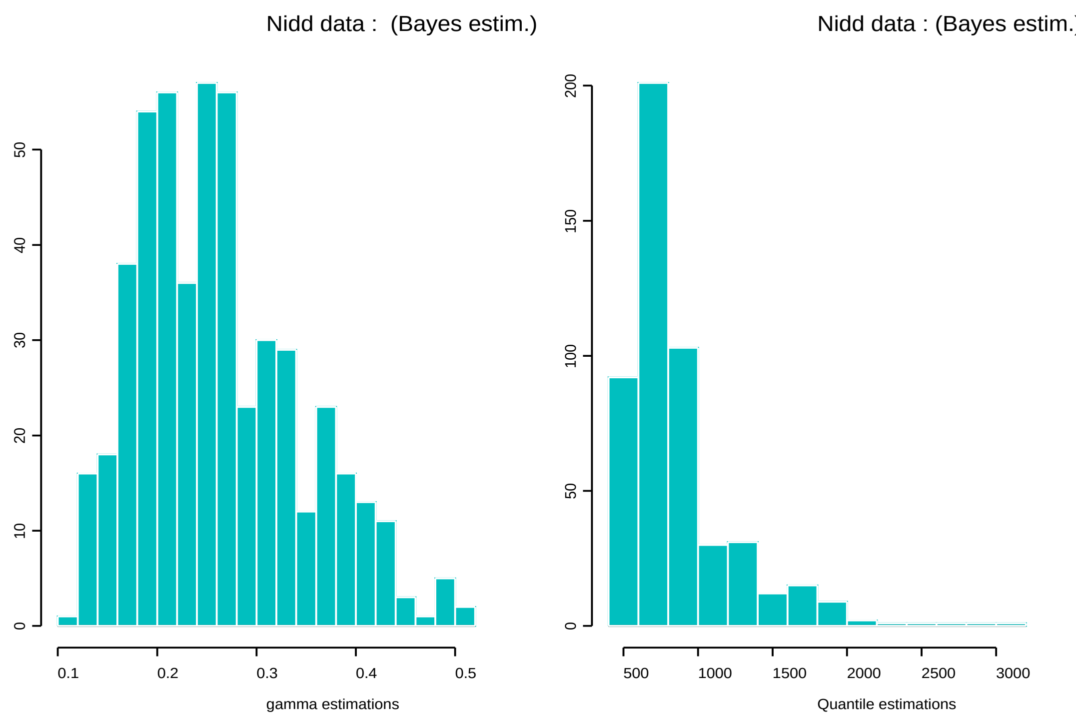

Figure 8: Histograms of 500 γ's and q_{1-p}'s simulated from the posterior distribution for the Nidd river data set, $k = 82$.

Analysis	ML point	Bayes-QC Return level estimate	Uniform Bayes Credibility Intervals	Bayes-QC Credibility Intervals
50-year return level				
$u = 100$	305	374	[210, 775]	[266, 672]
$u = 120$	280	403	[215, 850]	[304, 690]
100-year return level				
$u = 100$	340	457	[220, 935]	[306, 911]
$u = 120$	307	499	[225, 940]	[354, 961]

Table 1: Uniform and quasi-conjugate Bayesian 95 % credibility intervals for 50-year and 100-year return levels. Nidd river data.

4.2 Fire reinsurance data and net premium estimation

In the excess-of-loss (XL) reinsurance agreements, the re-insurer pays only for excesses over a high value u of the individual claim sizes. The *net premium* is the expectation of the total claim amounts that the re-insurer will pay during the future period $[0, T]$: $\mathbb{E}(S_{N_T}) = \sum_{i=1}^{N_T} Y_i$, where N_T is the random number of claims exceeding u during $[0, T]$ and $Y_1, Y_2, \ldots$ are the amounts of excesses above u. If the Y_i's are modeled by a GPD$(\gamma,\ \sigma)$ and the exceedance arrival process is modeled by a homogeneous Poisson process with annual rate λ, then the net premium over the coming year is approximated by $\mathbb{E}(S_{N_1}) \simeq \mathbb{E}(N_1)\,\mathbb{E}(Y_i) = \lambda\sigma/(1-\gamma)$.

Reiss and Thomas (1999, 2001) estimate the net premium of the Norwegian fire claims reinsurance by analysing a data set (see Table 2) of large Norwegian fire claims between 1983 and 1992. These data were initially presented by Schnieper (1993). They use a Bayesian inference method for the GPD parameters and assume that the exceedance process is Poisson. Independent gamma priors are used for λ and α and an inverse-gamma prior, with parameters $(a,\ b)$, is chosen for β. Posterior means of λ, α and β are computed using Monte Carlo numerical approximations of integrals. Table 3 compares estimates of $(\gamma,\ \sigma)$ and the net premium $\lambda\sigma/(1-\gamma)$ obtained by our quasi-conjugate approach with those obtained by Reiss and Thomas for ML and Bayesian estimates with different values of the hyperparameters a and b of the inverse-gamma. Our approach has the advantage of indicating the precision of the estimates.

5 Discussion

The proposed quasi-conjugate Bayes approach has many advantages when compared to standard GPD parameters and extreme quantiles estimators:

year	claim sizes (in millions)	year	claim size (in millions)
1983	42.719		23.208
1984	105.860	1990	37.772
1986	29.172		34.126
	22.654		27.990
1987	61.992	1992	53.472
	35.000		36.269
1988	26.891		31.088
1989	25.590		25.907
	24.130		

Table 2: Norwegian fire claims sizes over 22.0 millions NKr from 1983 to 1992, from Schnieper (1993).

Analysis	$\widehat{\gamma}$	$\widehat{\sigma}$	Net premium	
			Point estimate	90 % credib. interv.
ML for GPD	0.254	11.948	27.23	
Bayes (Reiss and Thomas)				
Inv.-Gamma($a = 2$, $b = 2$)	0.288	11.658	27.83	
Inv.-Gamma($a = 4$, $b = 6$)	0.274	11.814	27.66	
Quasi-conjugate Bayes approach	0.384	10.332	30.03	[17.09, 84.39]

Table 3: Bayesian estimates of γ, σ and the XL net premium for Norwegian fire reinsurance data.

– it can incorporate experts prior information and improve estimation of extreme events even when data are scarce,
– it provides Bayes credibility intervals assessing the quality of the extreme events estimation,
– it often gives estimators with week dependence on the number k of used excesses,
– the variances of the non-informative quasi-conjugate Bayes estimators compares well to the variances of the standard estimators. This suggests that informative quasi-conjugate Bayes estimators will give very accurate extreme quantile estimators, this point will be illustrated in a forthcoming paper.

We deeply describe the proposed quasi-conjugate Bayes approach for the most frequent case of Fréchet Maximum Domain of Attraction where tails are heavy ($\gamma > 0$), the case of Gumbel's Maximum Domain of Attraction ($\gamma = 0$) is analytically discussed in Appendix A.

Future work is needed to extend this approach to the general case where the user has no prior idea on γ.

The present paper is the first of a series of papers devoted to various developments of the Bayesian inference methodology that we introduced here. In the next two ones we will study: 1. How to determine and compute hyperparameters in a hierarchical structure setting based on the quasi-conjugate class defined here to take into account realistic expert prior information on extreme events ; 2. How to adapt the Hastings-Metropolis step of our Gibbs sampler on the basis of Worms' (2001) correction formulas to obtain bias-corrected results. This point is suggested by recent work based on the second-order theory of regularly varying functions on bias-corrected semiparametric estimators of γ and extreme quantiles. We think that a bias-corrected version of our Bayesian methodology will significantly improve estimation of extreme quantiles and probability tails.

Finally, note that it would be possible to include a Poisson parameter for the stream of exceedances as in Reiss and Thomas (2000). Also, spatial quantile estimation and multivariate or time-series extensions based on our approach are natural and very promising.

Appendix A

We present here a brief account of the simple case where Bayesian inference is made for exponential distributions, rather than GPD's with both parameters unknown. This simple setup is of interest since it is the Bayesian analogue of the Exponential Tail (ET) method (Breiman *et al.* 1990), and all calculations lead to explicit analytical formulas.

Set $\lambda = 1/\sigma$ and $f_\lambda(y) = \lambda\, e^{-\lambda y}\, \mathbf{1}_{y>0}$. Denote by $\pi_{a,\,b}$ the prior Gamma$(a,\, b)$ density $(a,\, b > 0)$. The posterior density $\pi_{a,\,b}(\lambda|\, y_1, \ldots, y_k)$ is Gamma$(a+k,\, b+S_k)$, where $S_k = \sum_{i=1}^{k} y_j$. Expert information is reflected in the choice of a and b. The corresponding posterior predictive distribution is GPD $(a+k,\, b+S_k)$, with $\gamma_{\text{pred}} = 1/(a+k)$ and $\sigma_{\text{pred}} = (b+S_k)/(a+k)$. Our first estimate of q_{1-p} is based on the posterior mean $\widehat{\lambda}_{\text{bayes}} = (a+k)/(b+S_k)$ of λ:

$$\widehat{q}_{1-p,\ \text{bayes}} \;=\; u \;+\; \frac{b+S_k}{a+k}\ln\left(\frac{k}{np}\right).$$

Our second estimate is based on the posterior predictive distribution:

$$\widehat{q}_{1-p,\ \text{pred}} \;=\; u \;+\; (b+S_k)\left[\left(\frac{k}{np}\right)^{1/(a+k)} -\, 1\right].$$

Our third estimate is based on the transformed posterior distribution. Since the posterior on $\sigma = 1/\lambda$ is an inverse-gamma distribution with density

$$\frac{(b+S_k)^{a+k}}{\Gamma\,(a+k)\ \sigma^{a+k+1}}\ \exp\left(-\,\frac{b+S_k}{\sigma}\right)\mathbf{1}_{\sigma>0}$$

whose mean is $(b+S_k)/(a+k-1)$ and variance is $(b+S_k)^2/(a+k-1)^2(a+k-2)$, the corresponding distribution of $u+\sigma \ln(k/np)$ has a similar shape, and has mean

$$\widehat{q}_{1-p,\,\text{post}} \;=\; u \;+\; \frac{b+S_k}{a+k-1} \ln\left(\frac{k}{np}\right)$$

and standard deviation

$$\frac{b+S_k}{(a+k-1)\sqrt{a+k-2}} \ln\left(\frac{k}{np}\right).$$

For k large enough, $\widehat{q}_{1-p,\,\text{bayes}}$ is close to $\widehat{q}_{1-p,\,\text{post}}$ with respect to the standard deviation scale, which is of the order of $(b+S_k)(a+k)^{-3/2}\ln(k/np)$. On the contrary, a Taylor expansion shows that when $\ln(k/np)/(a+k)$ is not too large,

$$\widehat{q}_{1-p,\,\text{pred}} \;\approx\; u \;+\; \frac{b+S_k}{a+k} \ln\left(\frac{k}{np}\right)\left[1 \;+\; \frac{\ln\left(\frac{k}{np}\right)}{2(a+k)}\right].$$

The distance between $\widehat{q}_{1-p,\,\text{pred}}$ and each of the two other estimates can be significant, and $\widehat{q}_{1-p,\,\text{pred}}$ exhibits a positive bias with respect to the other estimates. We have observed a similar behavior when dealing with GPD's: This is the reason why we have discarded the analogous of $\widehat{q}_{1-p,\,\text{pred}}$ in that setting, and selected estimates of q_{1-p} based on its posterior distribution. Finally, based on the transformation of posterior distribution, approximative 90 % (for instance) credibility intervals can be obtained through a very rough normal approximation to inverse-gamma distributions:

$$\left[u \;+\; \frac{b+S_k}{a+k-1} \ln\left(\frac{k}{np}\right)\left\{1 \;\pm\; 1.6\,\frac{1}{\sqrt{a+k-2}}\right\}\right].$$

References

[1] Bacro, J., and Brito, M. Strong limiting behaviour of a simple tail pareto-index estimator. *Statist. Dec.* **3**, 133–143, (1993).

[2] Balkema, A., and de Hannn, L. Residual life time at a great age. *Annals of Probability*, **2**, 792–804, (1974).

[3] Beirlant, J., Dierckx, G., Goegebeur, Y. and Mattys, G. Tail index estimation and an exponential regression model *Extremes* **2**, 177–200, (1999).

[4] Beirlant, J., Dierckx, G., and Guillou, A. Estimation of the extreme value index and regression on generalized quantile plots. Preprint, Catholic University of Leuven, Catholic University of Leuven, Belgium., (2001).

[5] Beirlant, J., Dierckx, G., Guillou, A., and Stărică, C. On exponential representations of log-spacings of extreme order statistics *Extremes* **5**, 157–181, (2002).

[6] BEIRLANT, J., AND TEUGELS, J. Asymptotic normality of Hill's estimator. In *Extreme Value Theory* (1989), eds., J. Hüsler and R.D. Reiss, Proceedings of the Oberwolfach Conference, 1987. Springer-Verlag New York.

[7] BEIRLANT, J., VYNCKIER, P., AND TEUGELS, J.L. Tail Index Estimation, Pareto Quantile Plots, and Regression Diagnostics *J. Amer. Stat. Assoc.* **91**, 1659–1667, (1996).

[8] BREIMAN, L., STONE, C., AND KOOPERBERG, C. Robust confidence bounds for extreme upper quantiles. *J. Statist. Comput. Simul.* **37**, 127–149, (1990).

[9] CAERS, J., BEIRLANT, J., AND VYNCKIER, P. Bootstrap confidence intervals for tail indices. *Computational Statistics and Data Analysis* **26**, 259–277, (1998).

[10] CASTILLO, E., AND HADI, A. Fitting the Generalized Pareto Distribution to data. *J. Amer. Stat. Assoc.* **92**, 440, 1609–1620, (1997).

[11] CHAUVEAU D., AND DIEBOLT J. An automated stopping rule for MCMC convergence assessment. *Computational Statistics* **14**, 419–442, (1999).

[12] COLES, S. *An introduction to statistical modelling of extreme values.* Springer series in Statistics, 2001.

[13] COLES, S., AND POWELL, E. Bayesian methods in extreme value modelling: A review and new developments. *International Statistical Review* **64**, 1, 119–136, (1996).

[14] COLES, S., AND TAWN, J. A Bayesian analysis of extreme rainfall data. *Applied Statistics* **45**, 4, 463–478, (1996).

[15] DAMSLETH, E. Conjugate classes of Gamma distributions. *Scandinavian Journal of Statistics: Theory and Applications* **2**, 80–84, (1975).

[16] DAVISON, A., AND SMITH, R. Models for exceedances over high thresholds. *J. R. Statist. Soc. B* **52**, 3, 393–442, (1990).

[17] DEKKERS, A., EINMAHL J.H.J., AND DE HAAN, L. A moment Estimator for the index of an extreme-value distribution. *The Annals of Statistics* **17**, 4, 1833–1855, (1989).

[18] DIEBOLT, J., GUILLOU, A., AND WORMS, R. Comportement asymptotique des estimateurs des moments pondérés et approximation pénultième. *Technical report of the dpt. of Mathematics of the University of Marne-la-Vallée 5/2002* (2002).

[19] DUPUIS, D. Exceedances over high thresholds: A guide to threshold selection. *Extremes* **3**, 1, 251–261, (1998).

[20] EMBRECHTS, P., KLÜPPELBERG, C., AND MIKOSCH, T. *Modelling Extremal Events.* Springer, 1997.

[21] Garrido, M. *Modélisation des évènements rares et estimation des quantiles extrêmes, Méthodes de sélection de modèles pour les queues de distribution.* PhD thesis, Université Joseph Fourier de Grenoble (France), 2002.

[22] Grimshaw, S. Computing maximum likelihood estimates for the Generalized Pareto Distribution. *Technometrics* **35**, 2, May, 185–191, (1993).

[23] Haeusler, E., and Teugels, J. On asymptotic normality of Hill's estimator for the exponent of regular variation. *The Annals of Statistics* **13**, 2, 743–756, (1985).

[24] Hill, B. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, **3**, 5, 1163–1174, (1975).

[25] Hosking, J., and Wallis, J. Parameter and quantile estimation for the Generalized Pareto Distribution. *Technometrics* **29**, 3, August, 339–349, (1987).

[26] Pickands, J. Statistical inference using extreme order statistics. *The Annals of Statistics*, **3**, 119–131, (1975).

[27] Reiss, R., and Thomas, M. A new class of Bayesian estimators in paretian excess-of-loss reinsurance. *Astin Journal* **29**, 2, 339–349, (1999).

[28] Reiss, R., and Thomas, M. *Statistical Analysis of Extreme Values.* Birkhauser Verlag, 2001.

[29] Robert, C.P. *Discretization and MCMC Convergence Assessment.* Lecture Notes in Statistics **135**, Springer Verlag, 1998.

[30] Schnieper, R. Praktische erfahrungen mit grossschadenverteilungen. *Mitteil. Schweiz. Verein Versicherungsmath*, 149–165, (1993).

[31] Schultze, J., and Steinebach, J. On least squares estimates of an exponential tail coefficient. *Statist. Decisions* **14**, 353–372, (1996).

[32] Singh, V., and Guo, H. Parameter estimation for 2-parameter Generalized Pareto Distribution by POME. *Stochastic Hydrology and Hydraulics* **11**, 3, 211–227, (1997).

[33] Smith, R. Estimating tails of probability distributions. *The Annals of Statistics* **15**, 3, 1174–1207, (1987).

[34] Smith, R., and Naylor, J. A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *Appl. Statist.* **36**, 3, 358–369, (1987).

[35] Worms, R. Vitesse de convergence de l'approximation de Pareto Généralisée de la loi des excès. *Comptes-Rendus de l'Académie des Sciences*, série 1, Tome 333, 65–70, (2001).

Contents

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)

Unité de recherche INRIA Futurs : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399